



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2019

---

## Mechanistic insights into mRNA 3'-end processing

Kumar, Ananthanarayanan ; Clerici, Marcello ; Muckenfuss, Lena M ; Passmore, Lori A ; Jinek, Martin

**Abstract:** The polyadenosine (poly(A)) tail found on the 3'-end of almost all eukaryotic mRNAs is important for mRNA stability and regulation of translation. mRNA 3'-end processing occurs co-transcriptionally and involves more than 20 proteins to specifically recognize the polyadenylation site, cleave the pre-mRNA, add a poly(A) tail, and trigger transcription termination. The polyadenylation site (PAS) defines the end of the 3'-untranslated region (3'-UTR) and, therefore, selection of the cleavage site is a critical event in regulating gene expression. Integrated structural biology approaches including biochemical reconstitution of multi-subunit complexes, cross-linking mass spectrometry, and structural analyses by X-ray crystallography and single-particle electron cryo-microscopy (cryoEM) have enabled recent progress in understanding the molecular mechanisms of the mRNA 3'-end processing machinery. Here, we describe new molecular insights into pre-mRNA recognition, cleavage and polyadenylation.

DOI: <https://doi.org/10.1016/j.sbi.2019.08.001>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-181146>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Kumar, Ananthanarayanan; Clerici, Marcello; Muckenfuss, Lena M; Passmore, Lori A; Jinek, Martin (2019). Mechanistic insights into mRNA 3'-end processing. *Current Opinion in Structural Biology*, 59:143-150.

DOI: <https://doi.org/10.1016/j.sbi.2019.08.001>



# Mechanistic insights into mRNA 3'-end processing

Ananthanarayanan Kumar<sup>1,3</sup>, Marcello Clerici<sup>2,3</sup>,  
 Lena M Muckenfuss<sup>2,3</sup>, Lori A Passmore<sup>1</sup> and Martin Jinek<sup>2</sup>

The polyadenosine (poly(A)) tail found on the 3'-end of almost all eukaryotic mRNAs is important for mRNA stability and regulation of translation. mRNA 3'-end processing occurs co-transcriptionally and involves more than 20 proteins to specifically recognize the polyadenylation site, cleave the pre-mRNA, add a poly(A) tail, and trigger transcription termination. The polyadenylation site (PAS) defines the end of the 3'-untranslated region (3'-UTR) and, therefore, selection of the cleavage site is a critical event in regulating gene expression. Integrated structural biology approaches including biochemical reconstitution of multi-subunit complexes, cross-linking mass spectrometry, and structural analyses by X-ray crystallography and single-particle electron cryo-microscopy (cryoEM) have enabled recent progress in understanding the molecular mechanisms of the mRNA 3'-end processing machinery. Here, we describe new molecular insights into pre-mRNA recognition, cleavage and polyadenylation.

## Addresses

<sup>1</sup> MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

<sup>2</sup> Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Corresponding authors:

Passmore, Lori A ([passmore@mrc-lmb.cam.ac.uk](mailto:passmore@mrc-lmb.cam.ac.uk)),

Jinek, Martin ([jinek@bioc.uzh.ch](mailto:jinek@bioc.uzh.ch))

<sup>3</sup> Equal contribution.

**Current Opinion in Structural Biology** 2019, **59**:143–150

This review comes from a themed issue on **Protein nucleic acid interactions**

Edited by **Frédéric H-T Allain** and **Martin Jinek**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 6th September 2019

<https://doi.org/10.1016/j.sbi.2019.08.001>

0959-440X/© 2019 MRC Laboratory of Molecular Biology. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Most eukaryotic pre-mRNAs are capped at their 5'-end, spliced at intronic sites, and polyadenylated at their 3'-end before they are exported from the nucleus as mature mRNAs. Each of these modifications is carried out by a set of conserved and highly regulated multi-protein complexes. The 3'-end processing machinery co-transcriptionally monitors nascent transcripts for specific sequences (Figure 1) and, upon recognition of the

polyadenylation site (PAS), cleaves the pre-mRNA and adds a poly(A) tail to the newly generated 3'-end [1]. The 3'-end processing machinery also triggers transcription termination. To co-ordinate these functions, it contains three different enzymatic activities — endonuclease, poly (A) polymerase, and protein phosphatase. Since the cleavage event defines the 3'-end of the transcript, and consequently the 3'-untranslated region (3'-UTR) of the future mRNA, understanding how RNA is specifically recognized is of key importance.

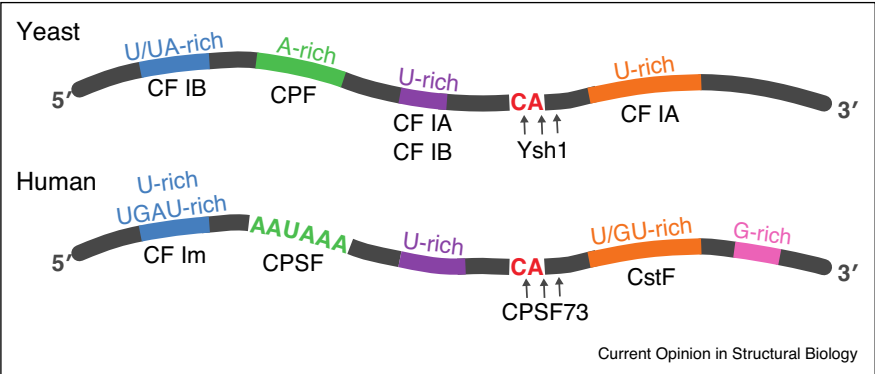
The large number of protein factors involved and the dynamic nature of their interactions pose challenges in understanding the molecular mechanisms of mRNA 3'-end biogenesis. Until very recently, high-resolution structures were available for only a few of the proteins involved [2,3]. Ysh1/CPSF73 had been identified as the 3'-endonuclease and its crystal structure was determined [4]. Crystal structures were also available for the poly(A) polymerase Pap1/PAP [5,6]. Together, these structures provided insights into the enzymatic mechanisms but they did not explain how pre-mRNAs are specifically recognized and how the different enzymatic activities are coordinated.

Recent developments in structure determination using electron cryo-microscopy (cryoEM), native mass spectrometry and cross-linking mass spectrometry (XL-MS), combined with next-generation sequencing-based functional methods, have facilitated progress in understanding the molecular mechanisms of pre-mRNA processing. In this review, we outline molecular details of the core machinery that mediates pre-mRNA cleavage and polyadenylation, focusing on insights obtained within the last two years. Structures and mechanisms of other components of the mRNA 3'-end processing machinery have been covered extensively by other reviews [2,3].

## Composition and functions of 3'-end processing machinery

Specific and efficient mRNA 3'-end processing is coordinated by the concerted action of a large number of conserved proteins (Table 1). Many components were first identified by cell fractionation studies, both from yeast and human extracts [1], and proteomic studies further defined the components of affinity-purified native complexes from *Saccharomyces cerevisiae* [7–10]. The yeast machinery comprises three factors: cleavage and polyadenylation factor (CPF), which contains the enzymatic

Figure 1



Polyadenylation signals and RNA recognition sites of the 3'-end processing machinery in yeast and human pre-mRNAs. Protein complexes that are proposed to recognize specific cis-acting elements within the polyadenylation signal are listed below their consensus sequences. Arrows indicate cleavage positions.

activities, cleavage factor (CF) IA and CF IB [1,11]. More recently, affinity-purification of an endogenous, intact CPF complex from yeast enabled thorough analysis of its composition, stoichiometry, and subunit assembly [12<sup>••</sup>]. Specifically, native mass spectrometry defined the protein–protein interaction network within the

complex, revealing that the CPF subunits are assembled into three modules, each based on one of the enzymatic activities: endonuclease, polymerase or phosphatase (Table 1). The interaction map of yeast CPF subunits enabled-specific cleavage and polyadenylation to be reconstituted from complexes of purified recombinant

**Table 1**  
**Yeast and human mRNA 3'-end processing machinery**

Yeast complex	Module	Yeast protein <sup>a</sup>	M.W. (kDa)	Human Protein <sup>a</sup>	M.W. (kDa)	Human complex	% Sequence identity (similarity)	Proposed role
CPF	Polymerase <sup>b</sup>	<b>Pap1</b>	65	<b>PAP</b>	83		40 (59)	<b>Poly(A) polymerase</b>
		Cft1/Yhh1	153	CPSF160/CPSF1	161		17 (34)	Scaffold
		Pfs2	53	WDR33	146		37 (57)	Scaffold, RNA binding
		Fip1	36	hFip1	67		22 (34)	Binds Pap1
		Yth1	25	CPSF30/CPSF4	30		36 (51)	RNA binding
	Nuclease <sup>b</sup>	Cft2/Ydh1	96	CPSF100/CPSF2	88	CPSF	22 (41)	(Pseudo-)endonuclease
		<b>Ysh1/Brr5</b>	88	<b>CPSF73/CPSF3</b>	77		44 (65)	<b>Endonuclease</b>
		Mpe1	50	RBBP6	202		28 (48)	RNA binding
		Pta1	89	Symplekin	141		17 (37)	Scaffold
		Ref2	60	n.d.	–		–	Regulates Glc7
	Phosphatase	Pti1	47	n.d.	–		–	Scaffold
		Swd2/Cps35	37	WDR82	35		34 (52)	Transcription termination
		<b>Glc7</b>	36	<b>PP1A</b>	38		85 (93)	<b>Phosphatase</b> of Pol II CTD
		<b>Ssu72</b>	24	<b>SSU72</b>	23		44 (65)	<b>Phosphatase</b> of Pol II CTD
CF IA		Rna14	78	CstF77/CSTF3	83		26 (42)	Stoichiometry of 2, scaffold
		Rna15	33	CstF64/CSTF2	61	CstF	27 (41)	Stoichiometry of 2, RNA binding
		n.d.		CstF50/CSTF1	48		–	
		Clp1	50	hClp1	48		27 (44)	RNA kinase in human
CF IB		Pcf11	72	hPcf11	173	CF IIm	21 (32)	Binds Pol II
		Hrp1	60	n.d.	–		–	RNA binding
		n.d.		CFIm25/CPSF5/NUDT21	26		–	RNA binding
		n.d.		CFIm68/CPSF6	59	CF Im	–	RNA binding
		n.d.		CFIm59/CPSF7	52		–	RNA binding

n.d., none detected.  
<sup>a</sup> Enzymes in bold.  
<sup>b</sup> Yeast polymerase and nuclease modules comprise the ‘Core CPF’ (CPF<sub>core</sub>).

yeast proteins [12<sup>••</sup>,13<sup>••</sup>]. *In vivo* transcriptome-wide mapping of yeast pre-mRNA biogenesis factors previously showed that CPF binds AU-rich elements near the cleavage site [14]. Still, isolated CPF is not substantially active without the accessory factors CF IA and CF IB, which contribute to RNA recognition and activation of the nuclease [3,15]. Both CF IA and CF IB bind-specific RNA sequences near the poly(A) site (Figure 1) [3,15].

The human mRNA 3'-end processing machinery comprises the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factors Im (CF Im) and IIm (CF IIm), and poly(A) polymerase (PAP) [3]. These factors include many orthologs of the yeast machinery (Table 1), but a fully active human complex has not yet been reconstituted from recombinant proteins. Whether the mammalian CPSF assembles in a similar modular fashion as CPF has not yet been confirmed but it appears to be functionally equivalent to the yeast core CPF complex (CPF<sub>core</sub>) comprising the nuclease and polymerase modules [13<sup>••</sup>,16<sup>•</sup>]. Some of the mammalian components, including WDR33 and Fip1, are much longer than the yeast subunits. Unlike in yeast, human PAP is not a stable component of the complex. A proteomic study of the human pre-mRNA 3'-end processing complex identified ~85 associated proteins [17<sup>•</sup>], but the functions of many of these have not been thoroughly investigated.

CPSF specifically recognizes a hexanucleotide AAUAAA motif within the polyadenylation signal (PAS), directing cleavage of the pre-mRNA 10–30 nucleotides downstream [16<sup>•</sup>,18<sup>•</sup>] (Figure 1). Transcriptome-wide studies of mammalian mRNA polyadenylation have shown that the AAUAAA PAS motif is highly conserved [19–21]. Recognition of upstream UGUA-containing sequences (USEs) and downstream G-rich and GU-rich sequences (DSEs) by the CF Im and CstF complexes, respectively, contributes to selection of the cleavage site and ensures efficient pre-mRNA recognition and cleavage [3].

### Molecular architecture of the polymerase module

Guided by the protein–protein interaction network map, a ~200-kDa recombinant four-subunit Cft1–Pfs2–Yth1–Fip1 complex of the polymerase module was recently analyzed using cryoEM. This resulted in a 3-D reconstruction comprising Cft1, Pfs2, and zinc finger (ZF) domains 1 and 2 of Yth1 [12<sup>••</sup>]. Cft1 contains three seven-bladed beta-propeller domains (BP1, BP2, BP3) followed by a C-terminal helical domain, and is intimately associated with Pfs2, a WD40-protein composed of a beta-propeller domain and an N-terminal protrusion that inserts into the cavity formed by Cft1 BP1 and BP3 (Figure 2a). The Cft1–Pfs2 interface is extensive, burying >4200 Å<sup>2</sup> surface area, and is highly conserved. The same interaction mode was observed in X-ray crystallographic and cryoEM structures of the orthologous human

CPSF160–WDR33 heterodimer [22<sup>••</sup>,23<sup>••</sup>,24<sup>••</sup>]. The triple beta-propeller domain architecture of Cft1/CPSF160 is structurally homologous to that of DNA Damage Binding protein 1 (DDB1) [25] and splicing factor SF3b subunit Rse1 [26] (SF3b130 in human) despite low sequence conservation (~15%). Interestingly, the interaction of DDB1 with its binding partner DDB2 has some similarities to the Cft1–Pfs2 and CPSF160–WDR33 interactions but the details of the subunit contacts are not conserved.

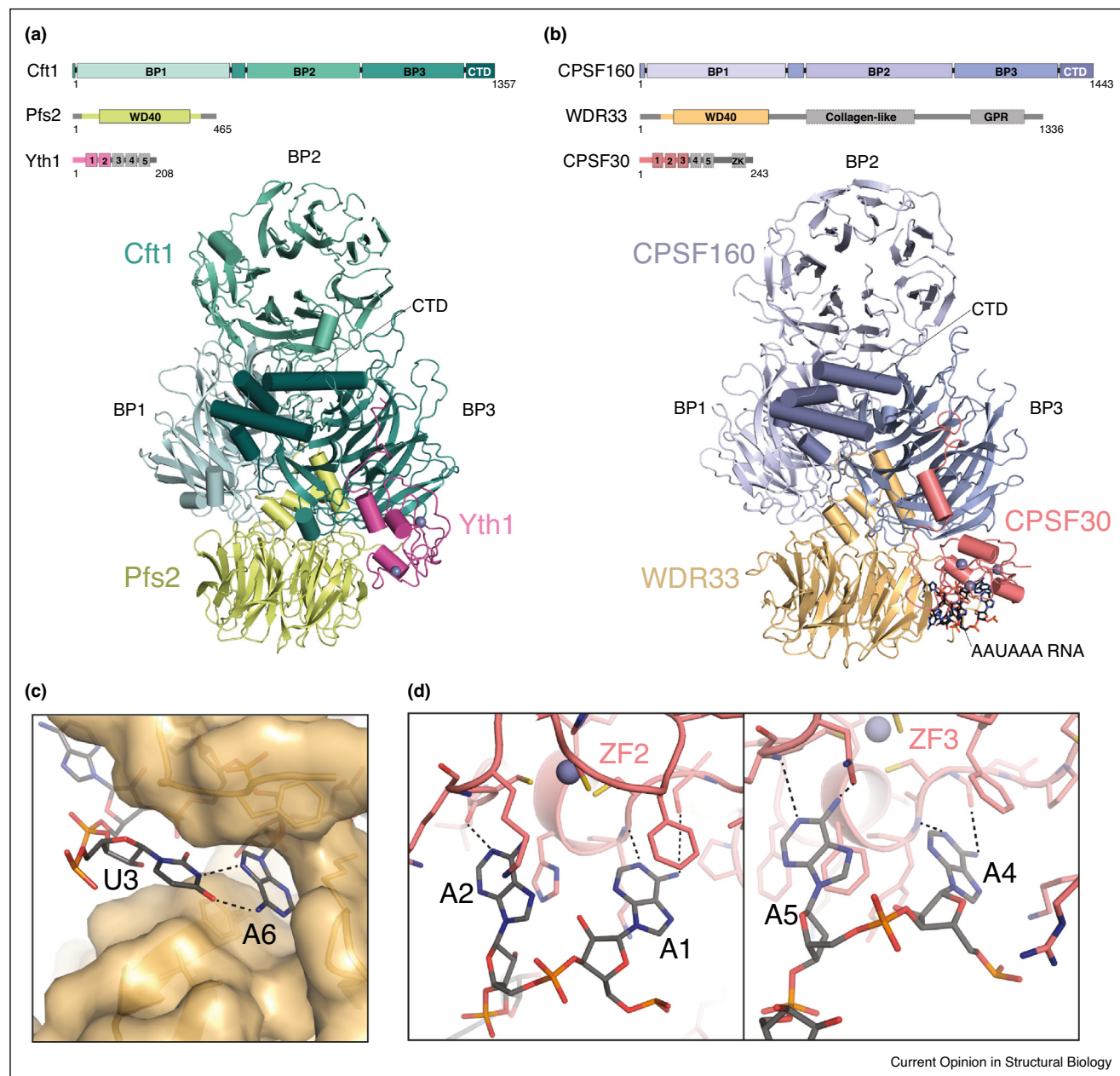
The Cft1–Pfs2 heterodimer constitutes a rigid core platform of the polymerase module which binds an N-terminal extended region and ZF1–2 of Yth1 (Figure 2a). ZF3–5 of Yth1 and the entirety of Fip1 are not observed in the cryo-EM reconstruction, indicating that they are flexibly tethered to Cft1–Pfs2, at least in the absence of other CPF subunits and/or RNA. Fip1, an intrinsically disordered protein, forms a physical connection between the C-terminal region of Yth1 and Pap1 [12<sup>••</sup>,27]. In an analogous manner, the human CPSF160–WDR33 heterodimer forms the structural scaffold of CPSF (Figure 2b), while CPSF30 bridges the interaction between CPSF160–WDR33 and Fip1, based on XL-MS analysis and co-precipitation experiments [22<sup>••</sup>]. Consistent with its absence in yeast Yth1, the CPSF30 zinc knuckle (ZK) domain is not required for complex assembly and its function is hitherto unknown [22<sup>••</sup>].

### Recognition of the polyadenylation signal

Similar to the yeast polymerase module, human CPSF160, WDR33, CPSF30 and Fip1 form a stable heterotetrameric core complex that recognizes the AAUAAA polyadenylation signal motif and recruits PAP [16<sup>•</sup>]. The molecular mechanism of PAS RNA recognition by the mammalian 3'-end processing machinery was recently revealed by two independently determined cryoEM structures of the human CPSF160–WDR33–CPSF30–Fip1 complex bound to AAUAAA-containing RNA. The overall architecture of the complex is highly similar to the yeast Cft1–Pfs2–Yth1–Fip1 structure. However, in the presence of the PAS RNA, the CPSF30 ZF3 domain becomes ordered (Figure 2b). PAS recognition is mediated by the CPSF30 ZF2 and ZF3 domains and WDR33 (Figure 2c and d) [23<sup>••</sup>,24<sup>••</sup>], as indicated by prior studies [16<sup>•</sup>,18<sup>•</sup>,27]. Although CPSF160 was previously implicated in PAS recognition [28], the RNA is not contacted by CPSF160 directly.

Recognition of the AAUAAA motif by CPSF30 involves bending of the RNA sugar-phosphate backbone into an S-shaped conformation, stabilized by extensive interactions with WDR33, in particular with its N-terminal extension that encircles the RNA. The kinked RNA conformation is further stabilized by an intramolecular Hoogsteen base pair formed between nucleotides U3 and A6 (Figure 2c). The adenosine dinucleotides in positions 1–2 and 4–5 of

Figure 2



Molecular architecture of the structural scaffolds of yeast CPF/human CPSF and mechanism of polyadenylation signal recognition.

**(a)** Structure of the yeast Cft1–Pfs2–Yth1 complex (PDB code 6EOJ). **(b)** Structure of the human CPSF160–WDR33–CPSF30 complex bound to AAUAAA PAS RNA (PDB code 6FUW). Protein regions not observed or present in the shown structures are indicated in grey in the domain architecture diagrams above each structure. **(c)** Specific recognition of the Hoogsteen base pair formed by U3 and A6 of the AAUAAA PAS by a hydrophobic cavity formed by the N-terminal extension of WDR33. **(d)** Molecular details of AA dinucleotide recognition by CPSF30 zinc finger (ZF) domains 2 (left panel) and 3 (right panel). Hydrogen-bonding interactions are depicted with dashed lines.

the AAUAAA motif are bound by the CPSF30 ZF2 and ZF3 domains, respectively (Figure 2d). Each nucleotide base is inserted into a pocket of the respective ZF domain and stacks with a conserved aromatic residue. Nucleotides A1, A4 and A5 are almost invariant in the PAS and

each adenine base is recognized by two base-specific hydrogen-bonding interactions with the N1 and N6-amino groups. In contrast, the adenine base of the more variable nucleotide A2 interacts with CPSF30 via a single hydrogen-bonding contact.



Contrary to the adenosine dinucleotides A1–A2 and A4–A5, the U3–A6 Hoogsteen base-pair is not recognized by base-specific interactions, but is sandwiched between two conserved phenylalanine residues which stabilize the base-pair with  $\pi$ – $\pi$  stacking interactions. The shape of the hydrophobic pocket formed by WDR33 to accommodate the U3–A6 pair is not compatible with other purine-pyrimidine combinations (Figure 2c). Thus, the intricate and specific network of molecular interactions established between CPSF30, WDR33 and all six nucleotide positions of the PAS provides a rationale for the widespread conservation of the AAUAAA motif revealed by transcriptome-wide mapping of mammalian mRNA polyadenylation sites [19–21]. In agreement with this, single-base substitutions in the PAS AAUAAA motif can result in a substantial reduction in the RNA binding affinity of the CPSF complex [22<sup>••</sup>] and in deficient mRNA processing in human diseases such as  $\alpha$ -thalassemia and  $\beta$ -thalassemia [29,30]. The structural insights suggest that non-canonical PAS motifs with one or more base substitutions [21] may function as weak polyadenylation sites: lower affinity of the non-canonical PAS motif for the CPSF complex may result in decreased use of that site. In the context of alternative polyadenylation, the weak affinity of non-canonical PAS motifs is likely compensated for by additional upstream and downstream *cis*-acting elements that enhance CPSF-binding under specific conditions or in response to specific signals.

In yeast, the polyadenylation signals (also termed positioning elements) are less well defined and typically contain degenerate A-rich motifs that often lack the U3–A6 Hoogsteen base pair nucleotides. In agreement with this, the U3–A6 binding pocket created by the N-terminal region of WDR33 is not conserved in yeast Pfs2. Conversely, the structure and sequence conservation of yeast Yth1 ZF2 and ZF3 domains, which specifically recognize A bases in higher eukaryotes, suggests that recognition of adenosine dinucleotides by Yth1 is conserved. Structural differences in the RNA recognition modes of the yeast and human machineries could explain the observations that yeast CPF<sub>core</sub> binds to model RNA substrates with much lower affinity than human CPSF [13<sup>••</sup>,22<sup>••</sup>].

### Mechanism of endonucleolytic cleavage

Ysh1/CPSF73, the endonuclease subunit of CPF/CPSF complexes, contains metallo- $\beta$ -lactamase (MBL) and  $\beta$ -CASP domains. The crystal structure of human CPSF73 revealed an active site at the junction of the two domains containing two coordinated zinc ions [4<sup>•</sup>] (Figure 3a). The geometry of the zinc ions, coordinating a hydroxide ion (the attacking nucleophile) and a sulfate ion (mimicking the scissile phosphate group in the RNA substrate) in the active site, suggests a possible catalytic mechanism (Figure 3a). However, the active site tunnel in this structure is very narrow and cannot accommodate

the RNA substrate. In agreement with this, purified human CPSF73 has only weak endonuclease activity *in vitro* [4<sup>•</sup>] and CPSF-dependent pre-mRNA cleavage has not been reconstituted *in vitro* to date. Together, these observations suggest that conformational activation of CPSF73 is required before pre-RNA cleavage. This may be important to prevent spurious nuclease activity and pre-mRNA misprocessing.

Recent biochemical reconstitution of the yeast pre-mRNA cleavage and polyadenylation machinery showed that the Ysh1 endonuclease subunit is only active when assembled into the ~500 kDa, 8-subunit 'CPF<sub>core</sub>' complex [13<sup>••</sup>]. Specific and efficient RNA cleavage also requires the presence of both CF IA and CF IB. Interestingly, the recombinant CPF<sub>core</sub> complex cleaved RNA substrates *in vitro* within a window of 3 nucleotides suggesting that it has positional accuracy, but does not have strict nucleotide specificity.

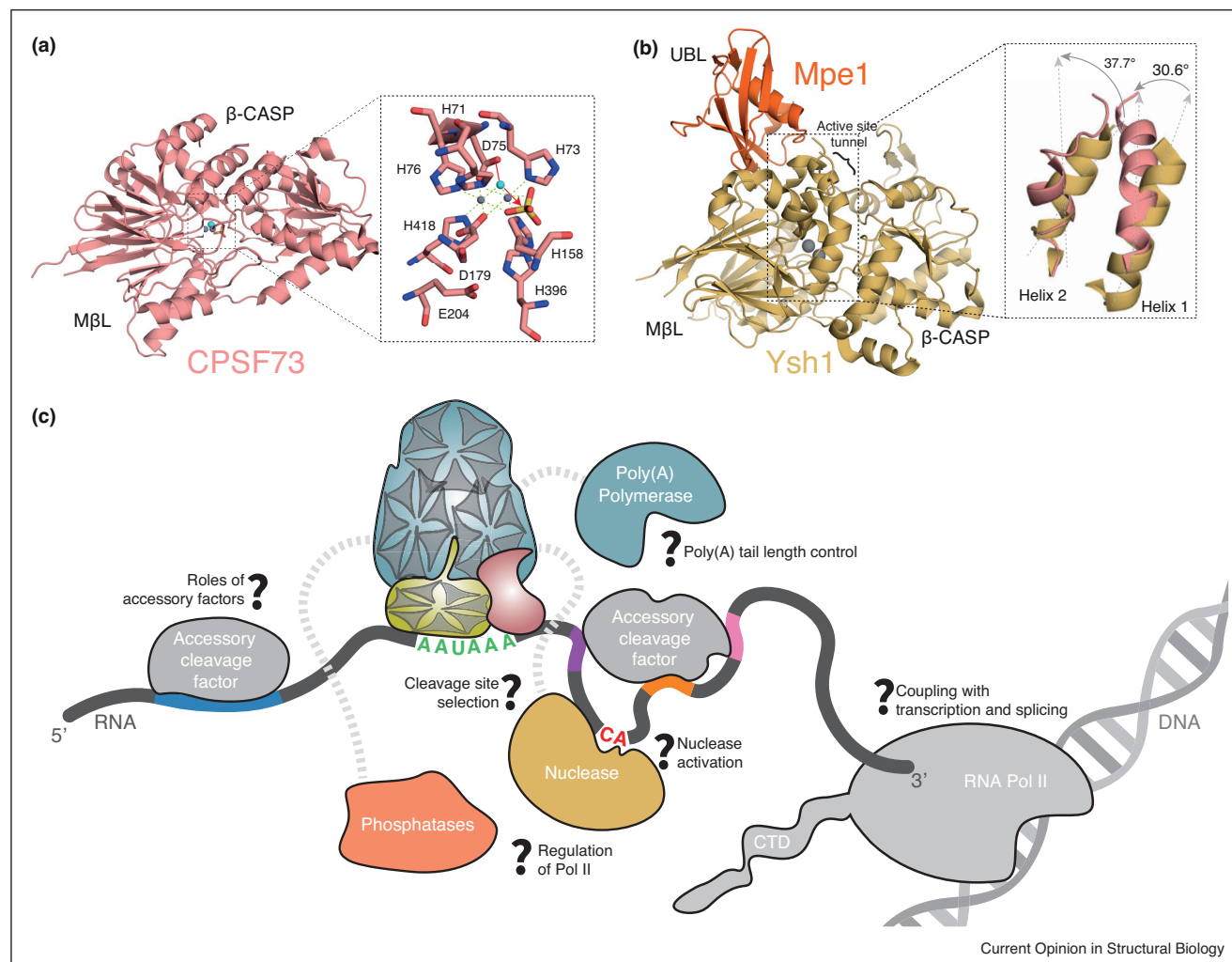
To gain insight into mechanisms that may prime Ysh1 for activation within CPF<sub>core</sub>, crystal and cryoEM structures of yeast Ysh1 in complex with the ubiquitin-like (UBL) domain of another CPF subunit Mpe1 (ortholog of human RBBP6) were determined [13<sup>••</sup>] (Figure 3b). These structures revealed an interaction interface burying ~900 Å<sup>2</sup> of surface area and involving hydrophobic, polar and electrostatic contacts. Two  $\alpha$ -helices within the Mpe1-bound Ysh1 MBL domain are shifted when compared to the CPSF73 structure (Figure 3b). This results in a slight opening of the cavity leading to the active site, but additional conformational changes must occur to accommodate the RNA substrate. The relatively high atomic B-factors of the  $\beta$ -CASP domain suggest that it may be mobile, allowing further opening of the active site cleft upon activation. Nevertheless, the precise mechanism underpinning this process and its dependence on other CPF/CPSF subunits and accessory polyadenylation factors still remains poorly understood.

### Conclusions and future challenges

Recent integrative structural studies have revealed the molecular architecture of the eukaryotic mRNA 3'-end processing machinery. Although these studies provide key insights into the assembly, RNA recognition and enzymatic activities of these factors, a number of questions concerning their molecular functions still remain unanswered (Figure 3c).

A critical aspect of the molecular mechanism of CPF/CPSF concerns the coupling of its RNA binding, endonuclease and poly(A) polymerase activities. A 3-D reconstruction of yeast CPF<sub>core</sub> obtained by negative-stain electron microscopy indicates that Cft1, Pfs2 and Yth1 of the polymerase module form a structural scaffold onto which Pap1 and the nuclease module are tethered [13<sup>••</sup>]. Conformational dynamics within CPF may serve to

Figure 3



Structures of the Ysh1/CPSF73 endonuclease and model for 3'-end processing.

**(a)** X-ray crystal structure of human CPSF73 showing the  $\beta$ -CASP and metallo- $\beta$ -lactamase (M $\beta$ L) domains (PDB code 2I7T). Zinc ions are shown as grey spheres and their coordination is indicated with dashed lines. In the active site, a hydroxide ion (shown as a blue sphere) is the attacking nucleophile and a sulfate molecule mimics the phosphate group of an RNA substrate. **(b)** Crystal structure of Ysh1 bound to ubiquitin-like domain (UBL) of Mpe1 (PDB code 6I1D). Inset shows superposition with the CPSF73 structure, revealing movement of two helices that could resemble a pre-activation state where the nuclease is primed for activation. Ysh1 is shown in gold, Mpe1 in orange and CPSF73 in pink. The active site tunnel lies behind the two helices formed by residues 95 to 111 (helix 1) and 122 to 136 (helix 2) of Ysh1 as indicated. **(c)** A model of mRNA 3'-end processing. The polymerase module hub of Cft1/CPSF160, Pfs2/WDR33 and Yth1/CPSF30 is shown with other subunits and factors surrounding it. Accessory cleavage factors are CF IA, CF IB, CstF, CF Im and CF Ilm. RNA is colored as in Figure 1. Questions remaining are indicated.

accommodate the variable distance between the polyadenylation signals and cleavage sites observed in different pre-mRNAs and might enable remodeling of the complex upon binding CF IA, CF IB and pre-mRNA. It is likely that similar mechanisms exist within the human complex. To understand the molecular basis of nuclease activation, future studies will focus on the structures of CPF/CPSF bound to accessory factors and additional *cis*-acting RNA elements.

Further studies will also be required to understand how the 3'-end processing machinery regulates transcription. It is known that a phosphatase subunit of yeast CPF (Ssu72) dephosphorylates Ser5 of the C-terminal domain (CTD) of Pol II to facilitate the switch from transcription initiation to elongation [31], while Glc7 dephosphorylates Tyr1 to promote transcription termination [32]. Structural studies will define how the phosphatase module interacts with CPF and whether the phosphatases are

mechanistically coupled to RNA recognition, cleavage and polyadenylation.

In mammalian cells, up to ~70% of genes have more than one PAS, giving rise to alternative mRNA isoforms that may encode different protein isoforms, or contain distinct 3'-UTRs conferring different stabilities and translational efficiencies [33]. The CPSF subunits Fip1 and RBBP6, as well as the accessory factors CstF and CF Im, have been implicated in the regulation of PAS selection [34–37]. Furthermore, interactions between the 3'-end processing factors and the pre-mRNA splicing machinery have also been shown to contribute to alternative polyadenylation [33,38]. The mechanistic basis for alternative polyadenylation remains to be determined.

Finally, mRNA poly(A) tails generally reach a defined length of ~60 As in yeast [39,40\*] and ~250 As in human [41]. Nuclear poly(A) binding proteins (yeast Nab2 and Pab1, and human PABPN1) play roles in regulating the activity and processivity of poly(A) polymerase, thus specifying the poly(A) tail length, but the molecular details of this process are unknown [42]. The recent structural and biochemical studies described here provide a foundation for future investigations that will focus on these aspects of mRNA 3'-end biogenesis and their regulation.

## Conflict of interest statement

Nothing declared.

## Acknowledgements

This work was supported by a Gates Cambridge scholarship (A.K.); the European Union's Horizon 2020 research and innovation programme (ERC Consolidator grant agreement No 725685) (L.A.P.); Medical Research Council grant MC\_U105192715 (L.A.P.); Boehringer Ingelheim Fonds PhD fellowship (to L.M.M.); Howard Hughes Medical Institute International Research Scholar Award (to M.J.) and the Bert N. and L. Kuggie Vallee Foundation (M.J.).

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Zhao J, Hyman L, Moore C: **Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis.** *Microbiol Mol Biol Rev* 1999, **63**:405-445.
2. Mandel CR, Bai Y, Tong L: **Protein factors in pre-mRNA 3'-end processing.** *Cell Mol Life Sci* 2008, **65**:1099-1122.
3. Xiang K, Tong L, Manley JL: **Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery.** *Mol Cell Biol* 2014, **34**:1894-1910.
4. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L: **Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease.** *Nature* 2006, **444**:953-956 This paper describes the crystal structures of human CPSF73 and yeast Cft2, and provides evidence that CPSF73 functions as the pre-mRNA endonuclease.
5. Bard J, Zhelkovsky AM, Helmling S, Earnest TN, Moore CL, Bohm A: **Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP.** *Science* 2000, **289**:1346-1349.
6. Martin G, Keller W, Doublet S: **Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP.** *EMBO J* 2000, **19**:4193-4203.
7. Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W: **Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination.** *EMBO J* 2002, **21**:4125-4135.
8. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM *et al.*: **Functional organization of the yeast proteome by systematic analysis of protein complexes.** *Nature* 2002, **415**:141-147.
9. Walsh EP, Lamont DJ, Beattie KA, Stark MJ: **Novel interactions of *Saccharomyces cerevisiae* type 1 protein phosphatase identified by single-step affinity purification and mass spectrometry.** *Biochemistry* 2002, **41**:2409-2420.
10. Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, Greenblatt J: **Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends.** *J Biol Chem* 2003, **278**:33000-33010.
11. Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S: **An interaction between U2AF 65 and CF I (m) links the splicing and 3' end processing machineries.** *EMBO J* 2006, **25**:4854-4864.
12. Casanal A, Kumar A, Hill CH, Easter AD, Emsley P, Degliesposti G, Gordiyenko Y, Santhanam B, Wolf J, Wiederhold K *et al.*: **Architecture of eukaryotic mRNA 3'-end processing machinery.** *Science* 2017, **358**:1056-1059 This work describes native MS analysis of the yeast CPF complex to determine its architecture, a cryoEM structure of the polymerase module, and biochemical analysis of polyadenylation.
13. Hill CH, Boreikaite V, Kumar A, Casanal A, Kubik P, Degliesposti G, Maslen S, Mariani A, von Loeffelholz O, Girbig M *et al.*: **Activation of the endonuclease that defines mRNA 3' ends requires incorporation into an 8-subunit core cleavage and polyadenylation factor complex.** *Mol Cell* 2019, **73**:1217-1231 Here, the authors reconstitute the cleavage activity of yeast CPF from recombinant proteins, and elucidate the architecture of CPF<sub>core</sub> using electron microscopy, cross-linking MS and hydrogen-deuterium exchange.
14. Baejen C, Torkler P, Gressel S, Essig K, Soding J, Cramer P: **Transcriptome maps of mRNA biogenesis factors define pre-mRNA recognition.** *Mol Cell* 2014, **55**:745-757.
15. Yang Q, Doublet S: **Structural biology of poly(A) site definition.** *Wiley Interdiscip Rev RNA* 2011, **2**:732-747.
16. Schonemann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E: **Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33.** *Genes Dev* 2014, **28**:2381-2393 This study reports the *in vitro* reconstitution of RNA polyadenylation by CPSF using recombinant proteins.
17. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL: **Molecular architecture of the human pre-mRNA 3' processing complex.** *Mol Cell* 2009, **33**:365-376 This study reports the isolation and MS analysis of the human 3'-end processing machinery.
18. Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR 3rd, Ule J, Manley JL, Shi Y: **CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing.** *Genes Dev* 2014, **28**:2370-2380 Here, the authors show that human CPSF recognizes the PAS RNA via its CPSF30 and WDR33 subunits.
19. Hu J, Lutz CS, Wilusz J, Tian B: **Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation.** *RNA* 2005, **11**:1485-1493.
20. Derti A, Garrett-Engle P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T: **A quantitative atlas of**



- polyadenylation in five mammals.** *Genome Res* 2012, **22**: 1173-1183.
21. Gruber AJ, Schmidt R, Gruber AR, Martin G, Ghosh S, Belmadani M, Keller W, Zavolan M: **A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation.** *Genome Res* 2016, **26**:1145-1159.
  22. Clerici M, Faini M, Aebersold R, Jinek M: **Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex.** *eLife* 2017, **6**This work describes the molecular architecture of the core of the human CPSF complex by cross-linking MS, biochemical analysis, and a crystal structure of the CPSF160-WDR33 heterodimer.
  23. Clerici M, Faini M, Muckenfuss LM, Aebersold R, Jinek M: **Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex.** *Nat Struct Mol Biol* 2018, **25**:135-138This paper describes the cryoEM structure of human CPSF160-WDR33-CPSF30 bound to PAS RNA, revealing the molecular mechanism of PAS recognition.
  24. Sun Y, Zhang Y, Hamilton K, Manley JL, Shi Y, Walz T, Tong L: **Molecular basis for the recognition of the human AAUAAA polyadenylation signal.** *Proc Natl Acad Sci U S A* 2017, **115**: E1419-E1428This paper describes the cryoEM structure of human CPSF160-WDR33-CPSF30 bound to PAS RNA, revealing the molecular mechanism of PAS recognition.
  25. Li T, Chen X, Garbutt KC, Zhou P, Zheng N: **Structure of DDB1 in complex with a paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase.** *Cell* 2006, **124**:105-117.
  26. Cretu C, Schmitzova J, Ponce-Salvatierra A, Dybkov O, De Laurentiis EI, Sharma K, Will CL, Urlaub H, Luhrmann R, Pena V: **Molecular architecture of SF3b and structural consequences of its cancer-related mutations.** *Mol Cell* 2016, **64**:307-319.
  27. Barabino SM, Ohnacker M, Keller W: **Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs.** *EMBO J* 2000, **19**:3778-3787.
  28. Keller W, Bienroth S, Lang KM, Christofori G: **Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA.** *EMBO J* 1991, **10**: 4241-4249.
  29. Higgs DR, Goodbourn SE, Lamb J, Clegg JB, Weatherall DJ, Proudfoot NJ: **Alpha-thalassaemia caused by a polyadenylation signal mutation.** *Nature* 1983, **306**:398-400.
  30. Orkin SH, Cheng TC, Antonarakis SE, Kazazian HH Jr: **Thalassemia due to a mutation in the cleavage-**  
**polyadenylation signal of the human beta-globin gene.** *EMBO J* 1985, **4**:453-456.
  31. Krishnamurthy P, Ross DD, Nakanishi T, Bailey-Dell K, Zhou S, Mercer KE, Sarkadi B, Sorrentino BP, Schuetz JD: **The stem cell marker Bcrp/ABCG2 enhances hypoxic cell survival through interactions with heme.** *J Biol Chem* 2004, **279**:24218-24225.
  32. Schrieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA: **RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7.** *Nat Struct Mol Biol* 2014, **21**:175-179.
  33. Tian B, Manley JL: **Alternative polyadenylation of mRNA precursors.** *Nat Rev Mol Cell Biol* 2017, **18**:18-30.
  34. Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi EA, Xie X, Wan J, Xing Y, Freudenberg JM *et al.*: **Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal.** *EMBO J* 2014, **33**:878-889.
  35. Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL *et al.*: **Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation.** *PLoS Genet* 2015, **11**:e1005166.
  36. Takagaki Y, Seipelt RL, Peterson ML, Manley JL: **The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation.** *Cell* 1996, **87**:941-952.
  37. Yao C, Choi EA, Weng L, Xie X, Wan J, Xing Y, Moresco JJ, Tu PG, Yates JR 3rd, Shi Y: **Overlapping and distinct functions of CstF64 and CstF64tau in mammalian mRNA 3' processing.** *RNA* 2013, **19**:1781-1790.
  38. Neve J, Patel R, Wang Z, Louey A, Furger AM: **Cleavage and polyadenylation: ending the message expands gene regulation.** *RNA Biol* 2017, **14**:865-890.
  39. McLaughlin CS, Warner JR, Edmonds M, Nakazato H, Vaughan MH: **Polyadenylic acid sequences in yeast messenger ribonucleic acid.** *J Biol Chem* 1973, **248**:1466-1471.
  40. Viphakone N, Voisinnet-Hakil F, Minvielle-Sebastia L: **Molecular dissection of mRNA poly(A) tail length control in yeast.** *Nucleic Acids Res* 2008, **36**:2418-2433A study of poly(A) tail length control by purified native yeast CPF.
  41. Wahle E: **Poly(A) tail length control is caused by termination of processive synthesis.** *J Biol Chem* 1995, **270**:2800-2808.
  42. Eckmann CR, Rammelt C, Wahle E: **Control of poly(A) tail length.** *Wiley Interdiscip Rev RNA* 2011, **2**:348-361.